Understanding Linguistic Accommodation in Code-Switched Human-Machine Dialogues

Tanmay Parekh\textsuperscript{1}, Emily P Ahn\textsuperscript{2}, Yulia Tsvetkov\textsuperscript{1}, Alan W Black\textsuperscript{1}

\textsuperscript{1} Carnegie Mellon University
\textsuperscript{2} University of Washington
Outline

• Introduction

• [C1] Bilingual Code-switching Dialogue System

• [C2] Data Collection

• [C3] Data Analysis

• Future Work
Introduction

Understanding **Linguistic Accommodation in Code-Switched Human-Machine Dialogues**
Introduction

One of the participants in the dialogue is a machine.

Understanding Linguistic Accommodation in Code-Switched Human-Machine Dialogues
Introduction

Usage of two or more languages within the same utterance or across utterances

I hope aapko mera talk and presentation acha lagega

[I hope you will like my talk and presentation]

Introduction

Dialogues between humans and machines wherein one or more participants are code-switching

Understanding Linguistic Accommodation in Code-Switched Human-Machine Dialogues

-Kya tumhare paas koi dost hai who like to eat mangoes?
-Nahi. Mere kisi friend ko aam pasand nahi
-Acha...mere 2 dost bank mein kaam karte hai and they like eating apple
Introduction

Phenomenon wherein people adjust their behaviors or speech styles to their conversational partners [2]

Understanding **Linguistic Accommodation** in Code-Switched Human-Machine Dialogues

- Pause Length
- Head Nodding
- Backchannels
- Posture
- Dialogue Acts
- Linguistic Style
- Utterance Length

Introduction

Understanding Linguistic Accommodation in Code-Switched Human-Machine Dialogues

Choice of language for specific words
Way of switching from one language to the other

Mere paas bhi 3 dost hai who like to eat apple

Acha...mere 2 dost bank mein kaam karte hai and they like eating apple
Motivation

Hey! How are you doing?

Hey! ¿Cómo estás?

Code-switching Chatbot

Make chatbots accommodate users

Hey! ¿Cómo estás?

I am doing good, ¿Y tú?

Engagement

Naturalness
Contributions

• [C1] Development of a generalized bilingual dialogue system that can be easily adapted to a new language pair

• [C2] New dataset - CommonDost - comprising of 439 Hindi-English human-machine conversations
Contributions

• [C3a] Analysis of accommodation of language style and language choice in the code-switched dialogues

• [C3b] Exploratory analysis comparing code-switching patterns across Spanish-English and Hindi-English
Outline

• Introduction

• [C1] Bilingual Code-switching Dialogue System

• [C2] Data Collection

• [C3] Data Analysis

• Future Work
Monolingual Dialogue System

Goal-Oriented Collaborative Dialogue System

Find your Mutual Friend

Time Remaining: 2:10

[02/09/18 00:17:18] <You entered the room.>
[02/09/18 00:17:19] Partner: hi
[02/09/18 00:17:28] Partner: I have 1 university of illinois at springfield, 1 radford university.
[02/09/18 00:17:44] You: sup
[02/09/18 00:17:50] Partner: Do you have any friends who like outdoor?
[02/09/18 00:17:57] You: I have a couple at radford

Your friends

<table>
<thead>
<tr>
<th>#</th>
<th>School</th>
<th>Time Preference</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Select</td>
<td>University of the Ozarks</td>
<td>afternoon</td>
<td>outdoor</td>
</tr>
<tr>
<td>Select</td>
<td>Christian Brothers University</td>
<td>afternoon</td>
<td>outdoor</td>
</tr>
<tr>
<td>Select</td>
<td>Radford University</td>
<td>afternoon</td>
<td>indoor</td>
</tr>
<tr>
<td>Select</td>
<td>North Carolina Wesleyan College</td>
<td>morning</td>
<td>outdoor</td>
</tr>
</tbody>
</table>

Bilingual Dialogue System

Bilingual Adaptations for Spanish-English [4]

Original Monolingual Dialogue System

Bilingual Adaptions

Bilingual Dialogue System: Issues

1. Reliance on a robust machine translation system (which can handle code-switching)
Bilingual Dialogue System: Issues

Linguistic rule-based adaptations vary largely based on language pair similarity.
Bilingual Dialogue System: Issues

Lack of normalization for writing code-switched text

- ek
- ik
- eik
- eek
- ikk

[one]

Code-switching

Latin Alphabet
- En
- Es
- Hi

Devanagiri Alphabet
- Ma
- Hi
Generalized Bilingual Dialogue System

Language Bias in KB
Generalized Bilingual Dialogue System

Handling Gender Markings

Favorite Fruit
Favorite Fruit
Profession
Profession
Hobby
Hobby
Time of Work
Time of Work
Gender
Gender

Pronouns
Pronouns

uska/uski
uska/uski
[his/her]
[his/her]

Verbs
Verbs

karta/karti
karta/karti
[he/she does]
[he/she does]

Gender Ratio
Gender Ratio
(M:F)
(M:F)

Carnegie Mellon University
Carnegie Mellon University
Language Technologies Institute
Language Technologies Institute

University of Washington
University of Washington
Generalized Bilingual Dialogue System

Dialogue Generation

Transformations

MT

English Text

Spanish Text

Spanish Text

Transformations

Phrase based MT

MT

English Text

Hindi Text

Hindi Text

Hinglish Text

Hinglish Text

Romanized Hindi Text

3
Generalized Bilingual Dialogue System

Spanglish Text → MT → English Text → NLU

Extracted Information

Hinglish Text → Rule-based NLU

Extracted Information

Natural Language Understanding
Outline

• Introduction

• [C1] Bilingual Code-switching Dialogue System

  • [C2] Data Collection

• [C3] Data Analysis

• Future Work
Data Collection

International Location
Screening Test (Audio Based)
Instructions in Hinglish
Conversational Task (Timer: 8 mins)
Post-Task Survey
CommonDost
Data Collection

Common dost koun hai?

Tum yaha ek aur online partner se baat karo. Tumhare aur us online user mein sirf ek common dost hai. Tumhe apne har dost ke baare mein kuch jaankari hai (jaise shauk, favourite phal, etc.). Iss jaankari se tumhe apna common dost dohondhna hai!

Instructions

- Please natural sentences ka upyog kijiye jitna ho sake.
  - Sahi: mere 3 dost bank mein kaam karte hai
  - Galat: 3 bank
- Seedha doston ki jaankari (kaam ki jagah ya samay, etc) mat likhiye. Pura sentence likhiye.
- Right side mein ek table mein tumhe apne doston ki jaankari milegi
- Tumhare partner ke paas bhi aisa hi ek table hai. Niche chat box mein partner se baat karke tumhe uske doston ki jaankari milegi. Tumhe us jaankari ka upyog karke common dost dohondhna hai
- Jab tumhe common dost mil jaaye, toh tum Select button dabake us dost ko chun sakhte ho. Agar tumne aur tumhare partner ne same dost ko chuna toh tum iss task mein safal ho jaoge
- Agar samay khatam ho bhi jaata hai par tumne achi koshish ki, tab bhi tumhe paise mileenge.
- Kripya dhyaan se chune. Yadi tumne kisi galat dost ko chuna, toh tumhe agle 10 second tak koi aur dost ko chunne ka mauka nahi milega. Uske baad tumhe phir se partner se baat karke dusra dost chunna hoga

Samay / Time: 7:51

[02/06/20 11:04:15] <You entered the room.>
[02/06/20 11:04:16] Partner: namaste

Enter your message here
Data Collection

Demographics

- 164 unique participants
- 72% Male
- 90% originate from Indian subcontinent
- 91% have a college degree

<table>
<thead>
<tr>
<th></th>
<th>Hinglish</th>
<th>Spanglish</th>
</tr>
</thead>
<tbody>
<tr>
<td># Dialogues</td>
<td>439</td>
<td>587</td>
</tr>
<tr>
<td># User Utterances</td>
<td>4,361</td>
<td>4,617</td>
</tr>
<tr>
<td># User Tokens</td>
<td>29,117</td>
<td>28,452</td>
</tr>
<tr>
<td>% Task Success</td>
<td>59%</td>
<td>64%</td>
</tr>
<tr>
<td>Avg dialogue length</td>
<td>9.93</td>
<td>7.9</td>
</tr>
<tr>
<td>Avg utterance length</td>
<td>6.68</td>
<td>6.2</td>
</tr>
<tr>
<td>EN vocab size</td>
<td>539</td>
<td>571</td>
</tr>
<tr>
<td>HI/SP vocab size</td>
<td>1,280</td>
<td>846</td>
</tr>
<tr>
<td>% EN utterances</td>
<td>19%</td>
<td>16%</td>
</tr>
<tr>
<td>% HI/SP utterances</td>
<td>34%</td>
<td>44%</td>
</tr>
<tr>
<td>% CS utterances</td>
<td>47%</td>
<td>39%</td>
</tr>
<tr>
<td>% CS dialogues</td>
<td>92%</td>
<td>70%</td>
</tr>
</tbody>
</table>
Data Collection

A: hey do you have any friends working at the zoo ya dost hai jise sona pasand hai [or friends who like sleeping]?

H: mere paas 2 dost hai jo zoo mei kaam karte hai aur unko photography ya drawing pasand hai respectively [I have 2 friends who work in the zoo and they like photography and drawing respectively]

A: toh [so] i have some female friends jinhe aam khana pasand hai [who like eating mango]

H: mere paas ek female friend hai jisko aam khana pasand hai aur usko dancing pasand hai [I have 1 female friend who likes eating mango and likes dancing]

A: i have two dost [friends] working at the machalighar [aquarium].

H: haaa [yes]

H: eek dosth hai [there is one friend]

A: i have one dost [friend] who likes sona [sleeping], one dost [friend] who likes machali pakadna [fishing] . . . aur tere [and yours]?

H: machali pakadna dho hai [two for fishing]
Outline

• Introduction

• [C1] Bilingual Code-switching Dialogue System

• [C2] Data Collection

• [C3] Data Analysis

• Future Work
Data Analysis: Background

Code-Switching Strategies

Insertional CS\(^5\)
Replace content words in matrix language with embedded language

Kya tumhare paas koi dost hai jise seb pasand hai?

Kya tumhare paas koi friend hai jise apple pasand hai?

[Do you have any friend who likes apple?]

Data Analysis: Background

**Code-Switching Strategies**

- Kya tumhare paas koi dost hai jise seb pasand hai?

**Alternational CS**

Switch from one language to another at a syntactic point

- Kya tumhare paas koi dost hai who likes apple?

[Do you have any friend who likes apple?]

---

Data Analysis: Background

Code-Switching Strategies

Informality in CS [6]

Addition of discourse markers

Kya tumhare paas koi dost hai jise seb pasand hai?

So kya tumhare paas koi dost hai jise seb pasand hai?

[Do you have any friend who likes apple?]

Data Analysis: Background

Linguistic Accommodation

Global Accommodation

Measuring speaker’s word usage and biasing it with all non-partners

\[ E_{a,b} = - \sum_{w \in V} |Pr_a(w) - Pr_b(w)| \]

\[ \text{ratio}(E_{a,b}, E_{a,np}) = \begin{cases} 
1 & E_{a,b} > E_{a,np} \\
0.5 & E_{a,b} = E_{a,np} \\
0 & E_{a,b} < E_{a,np} 
\end{cases} \]

\[ \text{global} = \frac{1}{|S|} \sum_{s \in S} \frac{1}{N_s} \sum_{np \in N_s} \text{ratio}(E_{s,p(s)}, E_{s,np}) \]

Data Analysis

We didn’t provide any explicit instructions to users to code-switch

92% CS dialogues

30% CS utterances even if agent is monolingual

47% CS user utterances

Our dataset elicits code-switching
Data Analysis

Task Success

<table>
<thead>
<tr>
<th></th>
<th>CS Agent</th>
<th>Mono Agent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Success</td>
<td>62%</td>
<td>48%</td>
</tr>
</tbody>
</table>

How non-native is your partner?

<table>
<thead>
<tr>
<th></th>
<th>CS Agent</th>
<th>Mono Agent</th>
<th>Random Agent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rating</td>
<td>2.62</td>
<td>2.92</td>
<td>3.11</td>
</tr>
</tbody>
</table>

Code-switching leads to better engagement
Data Analysis

- Informality improves dialogue quality

<table>
<thead>
<tr>
<th></th>
<th>Formal Agent</th>
<th>Informal Agent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg user utterance length</td>
<td>6.27</td>
<td>7.23</td>
</tr>
<tr>
<td># Avg Discourse Markers</td>
<td>1.87</td>
<td>2.44</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Formal Agent</th>
<th>Informal Agent</th>
</tr>
</thead>
<tbody>
<tr>
<td>How human-like was your partner?</td>
<td>3.54</td>
<td>3.99</td>
</tr>
</tbody>
</table>
Data Analysis

Accommodation of lexical items

If the agent uses any word for mentioning fruits in its utterance, will the user also use a word that refers to fruits in their utterance?

All

If the agent uses the *English* word for mentioning fruits in its utterance, will the user also use the *English* word that refers to fruits in their utterance?

Global Accommodation Score

<table>
<thead>
<tr>
<th>Language</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>0.79</td>
</tr>
<tr>
<td>English</td>
<td>0.648</td>
</tr>
<tr>
<td>Hindi</td>
<td>0.7</td>
</tr>
</tbody>
</table>

Agent’s language choice positively influences users’ language choice
Data Analysis

Confusion matrix for usage of user CS strategies compared to agent’s CS strategies

User uses a given strategy the most when the agent uses the same strategy

Agent’s CS strategy positively influences users’ CS strategy
Data Analysis

We ask users their other language of proficiency in the post-task survey

Higher Influence of English/Hindi in South/North Indian languages

South Indian Speakers

- Ins Hi->En
- Alt En->Hi

North Indian Speakers

- Ins En->Hi
- Alt Hi->En

Language Proficiency influences usage of CS strategies
Data Analysis

Equivalence Constraint: CS at point where syntactic rules are not violated

Word Order

Higher Alt in En-Es v/s En-Hi

Comparison of Spanglish and Hinglish

Outline

• Introduction

• [C1] Bilingual Code-switching Dialogue System

• [C2] Data Collection

• [C3] Data Analysis

• Future Work
Future Work

• Adaptation of the dialogue system to collect dialogue data to more CS languages

• Comparison of our findings about CS patterns and accommodation with human-human conversations

• Encourage further research for how linguistic and socio-political factors affect users’ CS distribution across language pairs
Team

Tanmay Parekh
Emily P Ahn

Yulia Tsvetkov
Alan W Black
Thank You
[dhanyawaad]